A Moment-Based Variational Approach to Tomographic Reconstruction*

Peyman Milanfar[†], W. Clem Karl, Alan S. Willsky

Laboratory for Information and Decision Systems
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, Massachusetts, 02139

| EDICS Number IP 2.3 Tomography |

December 12, 1993

Abstract

In this paper we describe a variational framework for the tomographic reconstruction of an image from the Maximum Likelihood 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.

The motivation for the approach presented herein stems from the idea that the moments of an image can be estimated directly from projections and used to focus the information available in the given projection data. This focusing of information in essence serves to control the degrees of freedom and regularize the reconstruction. Classical approaches to tomographic reconstruction are based on the idea of reconstructing every pixel value in the image directly; while in contrast, our approach uses a finite (usually relatively small) number of parameters (estimated moments) to compute reconstructions of the underlying image.

Furthermore, by focusing on the central 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, we can then adapt methods and concepts used for the latter problem in order to solve the former. In particular, 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. This solution has a statistical interpretation as the Maximum A-Posteriori estimate of the image given a divergence-based prior. We show that an iterative refinement of this algorithm leads to an efficient 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 to the classical Filtered Back-Projection algorithm.

1 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 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) while 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 [34], 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 also 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, an issue of considerable importance in problems with very noisy or sparse projection data. Another is the computational savings inherent to our approach. 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 other available algorithms operating on data limited in quality and quantity.

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 [22]. Furthermore, there has only been a little work in this area within the tomography community [28, 24, 16, 23, 8, 3]. On the other hand, the moment problem has been the subject of much work in the mathematics community for many years [1, 29]. However, while variants of variational/regularization methods developed here have been studied elsewhere in the literature [35, 34, 13, 21, 2], they have not been developed or investigated in the particular context of tomography of interest here. We also propose novel and efficient numerical techniques for solving this variational problem and study some of their properties and extensions. These algorithms may be of independent interest, although what we focus on here, for the most part, is the value of this approach for tomographic reconstruction from noisy and sparse data.

In Section 2 we present an optimal (Maximum Likelihood) algorithm for the estimation of the moments of an image from noisy measurements of its projections. In Section 3 we describe how the underlying image may be reconstructed from these estimated moments via regularization. Section 4 contains the explicit solution to this variational problem and here we also discuss those properties of this solution that make it attractive. In Section 5 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 6 contains our numerical simulation results including illustration of how prior information -in this case that provided by the standard filtered back-projection (FBP) solution- can be incorporated into our approach. Finally, in Section 7, we state our conclusions.

2 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 Figure 1.

The function $g(t, \theta) \in L^2([-1, 1] \times [0, 2\pi])$ [10] 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 [10], 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:

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

By considering $F(t) = e^{-it}$, the celebrated *Projection Slice Theorem* [11] 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 the case when $F(t) = P_k(t)$, where $P_k(t)$ is the k-th 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, Equation (2) relates the moments of the function f linearly to those of its Radon transform

4

g as we describe next.

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

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

Also, 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] $p+q \leq k$, which is a direct consequence of the consistency conditions for Radon transforms discussed in [10, 24]. 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 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 [18], we have used this decomposition of the Radon transform to derive new interpretations of classical

---

[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, while for k odd, it is a linear combination of λ_{pq} for $p+q = k$, $k-2$, $\cdots$, 3, 1.

reconstruction algorithms such as Filtered Back-Projection (FBP).

Suppose now that we are given noisy measurements of g at m distinct angles θ_1, θ_2, $\cdots$, θ_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] $-1 \leq t \leq 1$. 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 $\{P_k(t), \ k \geq 0\ \}$ 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 Maximum Likelihood (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 1 *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 $\lambda^{(j)}$, $0 \leq j < m$ of $f(x, y)$. This can be done using*

---

[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

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, 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

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

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, of course, optimal processing requires that *all* of these ML estimates be calculated. However, obviously 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, Figure 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.

3 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 which 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, 32, 33]. 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

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

The basic idea dates back to Kullback [14] and was later generalized by Csizar [6] and includes the principle of *Maximum Entropy* [12] 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 [31],

---

[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 $\widehat{\mathcal{L}}_N$

spectral analysis [30], image processing [36, 9] and recently tomography [28, 26, 7, 20, 3, 4]. In most of these applications, the general problem has often been posed as the following type of equality constrained optimization problem:

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

In particular, in the context of tomography, the weight functions $\phi_{i,j}(x,y)$ have been chosen as appropriate delta functions so that the constraints $\widehat{s}_{i,j}$ are the *noisy* measured values of the Radon transform $g(t_i, \theta_j)$ [28, 26, 7, 20]. 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 [13].

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

9

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 essentially *all* information of use in reconstruction.

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

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

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

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

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

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

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*[5] [13]. 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]Note that $D(\cdot, \cdot)$ is *not* a true metric since $D(f, f_0) \neq D(f_0, f)$.

4 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]. Also 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, the cost functional $J_{IDR}(f)$ can be written as

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

The cost functional $J_{IDR}(f)$ has a unique minimum due to its convex nature [14, 13]. Furthermore, a straightforward variational calculation analogous to ones in other I-divergence minimization problems [13, 32] (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

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

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.

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

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

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

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

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. This makes the IDR approach a computationally attractive one. Also, 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.

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

$$\widehat{f}_{k+1} = \arg\min_f J_k(f, \widehat{f}_k). \tag{29}$$

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

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

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. Secondly, 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 [29].

The idea of using iterative methods to solve divergence-based minimization problems has been considered in other contexts [6, 33, 3, 27]. Distinct features of our approach are the applications to tomography[6]

---

[6]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].

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, 33, 3]. In Appendix C we provide a convergence result for our specific context which in particular provides us with guidance on how the regularization parameter γ_k should be chosen at each iteration. Moreover, note that, assuming that γ_k is chosen to ensure convergence, then 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).

6   Numerical Examples

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 *per sample*.

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

14

In addition, to quantify the quality of the reconstructions, we define the *percent* Mean-Squared-Error (% MSE) as follows:

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

Example I: The first phantom to be reconstructed is a 64 by 64 grayscale image shown in the upper left corner of Figure 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 Gaussian white noise so as to produce an overall SNR of 4.35 dB per sample. In the lower left side of Figure 3, the Filtered Back-Projection (FBP) reconstruction is shown where a Butterworth filter of order 3 with cutoff frequency of 0.25 (normalized) was used. This choice of filter and cutoff 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 so as 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 3 iterations is shown in the lower right-hand side of Figure 3, while 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% while after only 3 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 Figure 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 inclusion of additional moments. This is illustrated in Figure 5 which shows the MSE versus the order of the highest order moment used in the reconstructions. Figure 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 percent MSE by only roughly 1 percent, 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 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 Figure 7, we show the reconstructions when a uniform prior is used. As we have pointed out previously, this corresponds to a 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. Figure 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: $f_0 = \arg\min_f \gamma_0 \int\int_D f - \log(f) dx dy + (\widehat{\mathcal{L}}_N - \mathcal{L}_N(f))^T \Sigma_N (\widehat{\mathcal{L}}_N - \mathcal{L}_N(f))$. As is apparent, in contrast to the maximum (Shannon) entropy solution in Figure 7, the Burg entropy solution is known to give "peaked" or "spikey" results [13]. It is interesting to contrast the It-IDR solutions in the upper right-hand corners of Figures 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 Figures 7 and 8) do *not* have high frequency noise, the It-IDR reconstructions in these cases also do not exhibit such noise. This in contrast to the FBP prior (in the lower left of Figure 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, of course, on the application.

Example II: The second phantom to be reconstructed is a 64 by 64 grayscale image shown in the upper left corner of Figure 9, 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 so as to produce an overall SNR of 4.35 dB per sample. In the lower left side, the Filtered Back-Projection (FBP) reconstruction is shown where a Butterworth filter of order 2 with cutoff 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 1 iterations (i.e. the IDR solution) is shown in the lower right-hand side of Figure 9, while 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.

The It-IDR algorithm performs well even when a much smaller number of projections is available. As shown in Figure 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.

7 Conclusions

In this paper we have shown how the tomographic reconstruction problem can naturally be decomposed into a two-step process whereby we first compute maximum likelihood estimates of the orthogonal moments of the underlying image directly from the projections, and then use these estimated moments to obtain a reconstruction of of the image. In particular, by 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 from noisy projection data which far exceed, in quality, those reconstructions produced by classical tomographic reconstruction techniques.

We demonstrated how our proposed algorithm provides an explicit mechanism for controlling the degrees of freedom in the reconstructions and hence resulting in better results. Also, in contrast to other divergence- (or entropy-) based algorithms which 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 [25, 15], or a preliminary parameterized reconstruction such as polygonal reconstructions [19], an estimate is obtained of the region of the plane where the object of interest is 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. 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 computationally efficient tomographic reconstruction for low signal-to-noise ratio scenarios and when the number of available projections may be small.

A Proof of Proposition 1

This result is most easily proved using the nonorthogonal geometric moments

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

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

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

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

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

where $\gamma_{k,j} = \frac{k!}{j!(k-j)!}$ 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 obviously 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 $\theta = \theta_i, \; \cdots, \; \theta_m$:

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

To see that this cannot happen for any such $p_N(\theta)$ for any N and m satisfying $N \le m-1$, we proceed by induction on both N and m. 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 $m = 1$ and *all* $N \le m-1$ (i.e. $N = 0$) and for any $m > 1$ and $N = 0$. Thus suppose by induction that the result is true for all $N \le m-1$ and for all $m \le p-1$ and also for all $N \le k-1$ for $m = p$; where $k \le p-1$. What we need to show is that it is also true for $N = k$, $m = p$. So take any nonzero $p_k(\theta)$. Note first that if $p_k(\pi/2) = 0$, then from (41) (with $N = k$), we see that

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

Therefore we have that

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

If one of the θ_j, say $\theta_p = \pi/2$, then what we want is that $p_{k-1}(\theta_j)$ can not vanish for all $j = 1, \cdots, p-1$; 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 $p_{k-1}(\theta_j)$ can not vanish for $j = 1, \cdots, p$, which is also part of the induction hypothesis.

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

$$p_k(\theta) = \cos^k(\theta)q_k(\theta) \qquad (44)$$

where

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

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 [17], 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 p values of θ_j, proving the result.

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 [14, 32, 6] that the unique solution to (33) has the form[7]

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

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 (namely 0). Now through (32), this implies 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

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

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

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

---

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

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 $< \cdot, \cdot >_{\Sigma_N}$

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

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

$$\widehat{f}_{k+1} = \arg\min_f \gamma_k D(f, \widehat{f}_k) + \|\mathcal{L}_N(f) - \widehat{\mathcal{L}}_N^{(c)}\|_{\Sigma_N}^2 + \|\widehat{\mathcal{L}}_N^{(i)}\|_{\Sigma_N}^2 \tag{52}$$

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).

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

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

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

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

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. i.e.

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

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

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

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

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

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

$$\text{Condition 2:} \quad \| \prod_{j=1}^{k} T_j \|_2 \leq c \alpha^k \tag{58}$$

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.* Hafner, New York, 1965.

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

[3] Charles L. Byrne. Iterative image reconstruction algorithms based on cross-entropy minimization. *IEEE Trans. on Image Processing*, 2(1):96–103, January 1993.

[4] A.T. Chinwalla and J. A. O'Sullivan. Image regularization using a divergence penalty method. In *Proceedings of the Johns Hopkins Conference on Information Science and Systems*, pages 30–33, Baltimore, MD, March 1993. Johns Hopkins University.

[5] S. Conte. *Elementary Numerical Analysis.* Mc Graw-Hill, 1965.

[6] I. Csiszar. I-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3:146–158, 1975.

[7] N.J. Dusaussoy and I.E. Abdou. The extended MENT algorithm: A maximum entropy type algorithm using prior knowledge for computerized tomography. *IEEE Trans. on Signal Processing*, 39(5):1164–1180, May 1991.

[8] M. Ein-Gal. *The shadow transformation: an approach to cross-sectional imaging.* PhD thesis, Stanford University, 1974.

[9] S.F. Gull and G.J. Daniell. Image reconstruction from incomplete and noisy data. *Nature*, 272:686–690, 1978.

[10] Sigurdur Helgason. *Radon Transform.* Birkhauser, Boston, 1980.

[11] G. T. Hermann. *Image Reconstruction From Projections.* Academic Press, New York, 1980.

[12] Edwin T. Jaynes. On the rationale of maximum entropy methods. *Proceedings of the IEEE*, 70(9):939–952, Sept 1982.

[13] Lee K. Jones and Charles L. Byrne. General entropy criteria for inverse problems, with applications to data compression, pattern classification, and cluster analysis. *IEEE Transaction on Information Theory*, 36(1), 1990.

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

[15] Avinash Lele. Convex set estimation from support line measurements. Master's thesis, MIT, Dept. of EECS, 1990.

[16] A. K. Louis. Picture restoration from projections in restricted range. *Math. Meth. in the Appl. Sci.*, 2:209–220, 1980.

[17] S. MacLane and G. Birkhoff. *Algebra.* Chelsea Publishers, New York, 1988.

[18] Peyman Milanfar. *Geometric Estimation and Reconstruction from Tomographic Data.* PhD thesis, M.I.T., Department of Electrical Engineering, June 1993.

[19] Peyman Milanfar, W. Clem Karl, and Alan S. Willsky. Reconstructing finitely parameterized objects from projections: A statistical view. *Submitted to CVGIP: Graphical Models and Image Processing*, August 1993.

[20] Geraldo Minerbo. MENT: A maximum entropy algorithm for reconstructing a source from projection data. *Computer Graphics and Image Processing*, 10:48–68, 1979.

[21] Frank Natterer. *Mathematics and Computer Science in Medical Imaging*, volume F39 of *NATO ASI Series*, chapter Regularization Techniques in Medical Imaging, pages 127–141. Springer Verlag, 1988.

[22] Miroslaw Pawlak. On the reconstruction aspects of moments descriptors. *IEEE Trans. Info. Theory*, 38(6):1698–1708, November 1992.

[23] A. Peres. Tomographic reconstruction from limited angular data. *J. Computer Assisted Tomography*, 3(6):800–803, 1979.

[24] Jerry L. Prince and Alan S. Willsky. Constrained sinogram restoration for limited-angle tomography. *Optical Engineering*, 29(5):535–544, May 1990.

[25] Jerry L. Prince and Alan S. Willsky. Reconstructing convex sets from support line measurements. *IEEE Trans. Pat. Anal. Mach. Intell.*, 12(4):377–389, 1990.

[26] M.L. Reis and N.C. Roberty. Maximum entropy algorithms for image reconstruction from projections. *Inverse Problems*, 8:623–644, 1992.

[27] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Society of America*, 1972.

[28] M. Ibrahim Sezan and Henry Stark. Incorporation of a priori moment information into signal recovery and synthesis problems. *Journal of Mathematical Analysis and Applications*, 122(172-186), 1987.

[29] J.A. Shohat and J.D. Tamarkin. *The problem of moments*. American Mathematical Society, New York, 1943.

[30] J.E. Shore. Minimum cross-entropy spectral analysis. *IEEE Trans. Acoust. Speech, Signal Processing*, ASSP-29(2):230–237, Apr. 1981.

[31] J.E. Shore and R.M. Gray. Minimum cross-entropy pattern classification. *IEEE Trans. on Pattern Anal. and Mach. Intell.*, PAMI-4(1):11–17, Jan. 1982.

[32] J.E. Shore and R.W. Johnson. Properties of cross-entropy minimization. *IEEE Trans. on Info. Theory*, IT-27(4), July 1981.

[33] D.L. Snyder, T.J. Shulz, and J.A. O'Sullivan. Deblurring subject to nonnegativity constraints. *IEEE Trans. on Signal Processing*, 40(5):1143–1150, May 1992.

[34] G. Talenti. Recovering a function from a finite number of moments. *Inverse Problems*, 3:501–517, 1987.

[35] A.N. Tikhonov and V.Y. Arsenine. *Solutions of Ill-posed Problems*. Winston/Wiley, Washington, D.C., 1977.

[36] S.J. Wernecke and Larry D'Addario. Maximum entropy image reconstruction. *IEEE Transactions on Computers*, C-26, 1977.
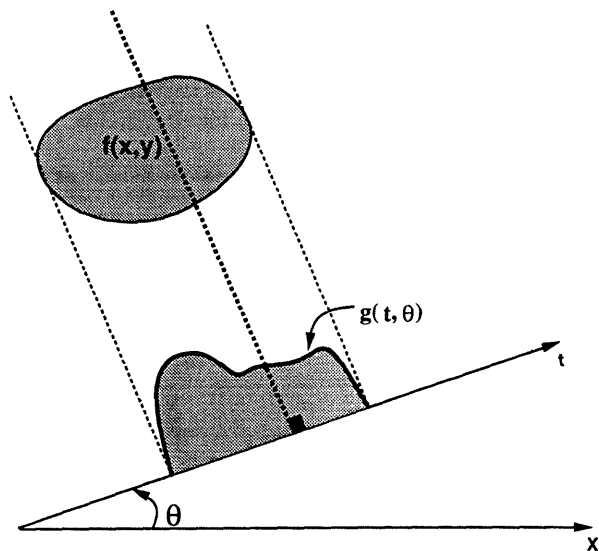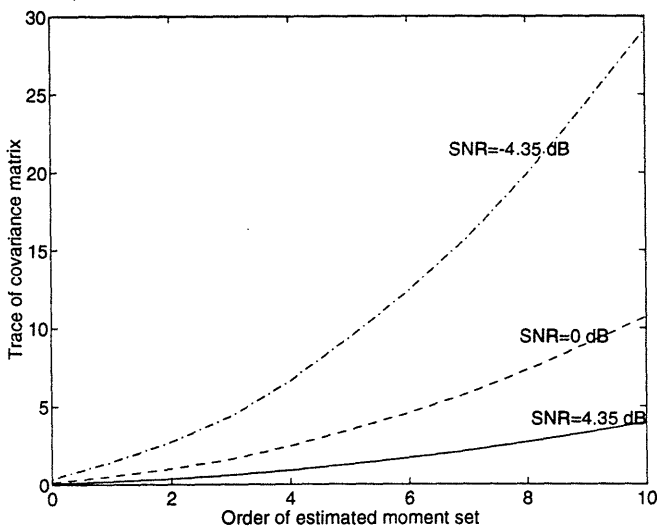
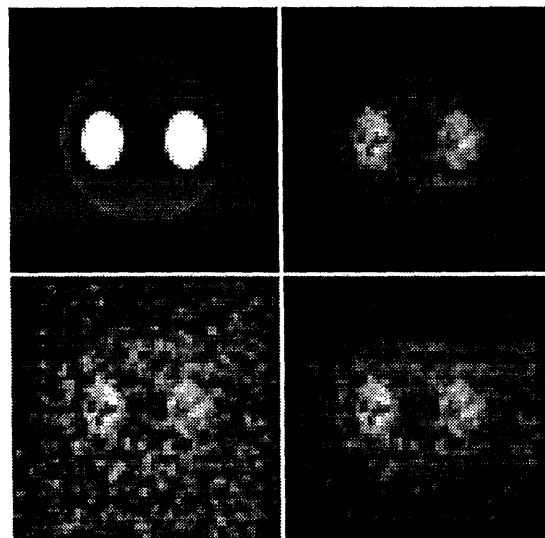Figure 1: The Radon transform

Figure 3: Counter-clockwise 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. w/ 64 samples per proj. SNR = 4.35 dB; moments up to order 8 used.
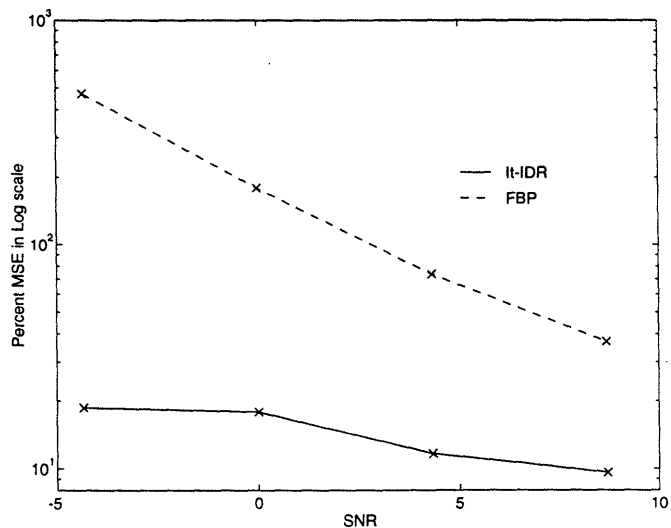

Figure 2: Trace of covariance matrix versus moment order up to order 10

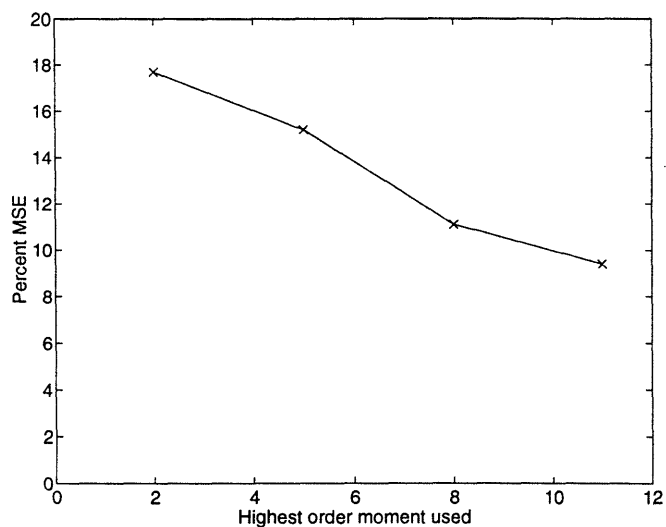Figure 4: MSE versus SNR (dB) in reconstructing the phantom of Figure 3: Moments up to order 8 used.

Figure 5: MSE versus number of moments used in It-IDR reconstruction of the phantom of Figure 3 with SNR=4.35 dB
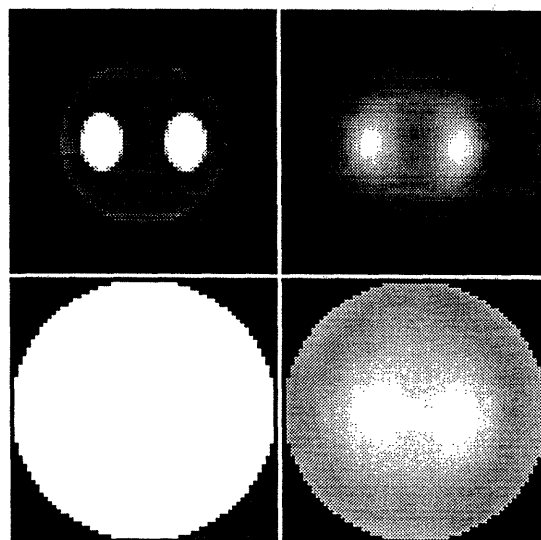
Figure 7: Counter-clockwise from upper left: Phantom, uniform initial estimate (% MSE=65.7), IDR solution (% MSE=55.9), It-IDR solution (% MSE=15.8). Data: 64 projections w/ 64 samples per projection at SNR= 4.35 dB; moments up to order 8 were used.
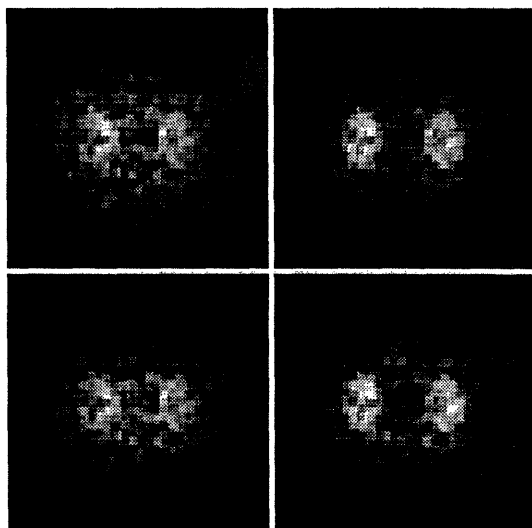

Figure 6: Counter-clockwise from upper left: reconstructions using moments up to order 2, 5, 8, and 11. Data: 64 projections w/64 samples per projection at SNR= 4.35 dB. Initial guess was based on FBP in every case.

Figure 9: Counter-clockwise from top left: Phantom, Initial Estimate from FBP, IDR reconstruction, Final It-IDR reconstruction (64 views, 64 samples per view, SNR =4.3 dB, moments up to order 10 used)
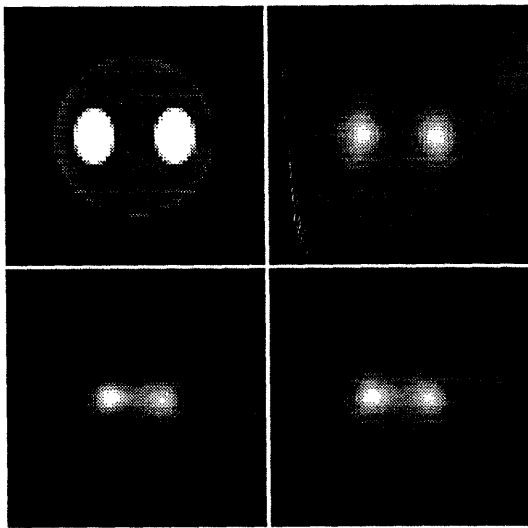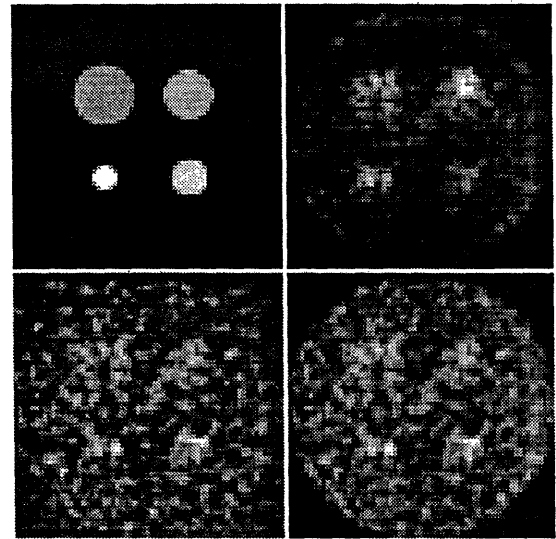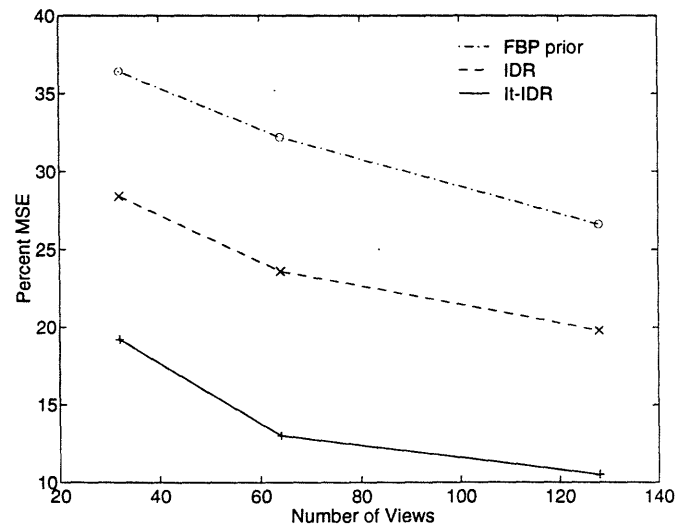

Figure 8: Counter-clockwise from upper left: Phantom, Initialization computed using Burg entropy (% MSE=39.8), IDR solution (% MSE=31), It-IDR solution (% MSE=10.3). Data: 64 projections, 64 samples per projection with SNR=4.35 dB; moments up to order 8 used.

Figure 10: MSE versus number of views for "adjusted" FBP prior, IDR solution and It-IDR solution, SNR=4.3 dB, moments up to order 10 used

29